PCIe® 5.0 PHY Logical

Dr. Debendra Das Sharma

Intel Fellow

Director of I/O Technology and Standards

Data Center Group, Intel Corporation

Disclaimer

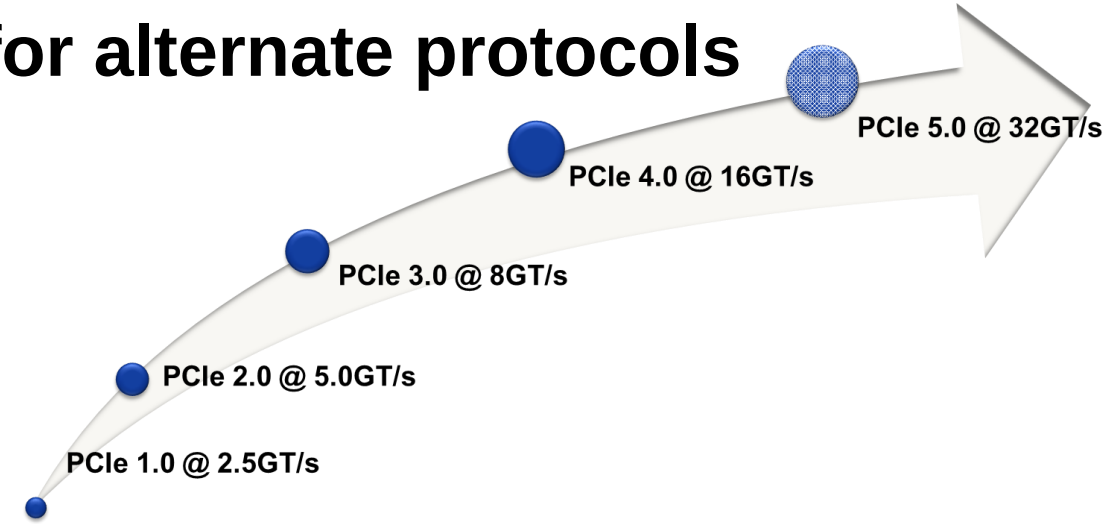
The information in this presentation refers to specifications still in the development process. This presentation reflects the current thinking of various PCI-SIG® workgroups, but all material is subject to change before the specifications are released.

Agenda

- **Background**
- **128b/130b Changes for PCIe® 5.0**
- **Transmitter Equalization and Training**
- **Precoding**
- **Alternate Protocol Negotiation**
- **Testability Features**
- **Configuration Registers for 32GT/s**
- **Summary**

PCIe 5.0 – Backwards Compatible with Prior Generations at 32GT/s

- **128b/130b encoding with minor changes to Ordered Sets (no changes to Data Blocks)**
- **Equalization flows similar to PCIe 4.0 with enhancements such as EQ bypass, precoding**
- **Retimers (up to 2) for channel extension**
- **Support for alternate protocols**



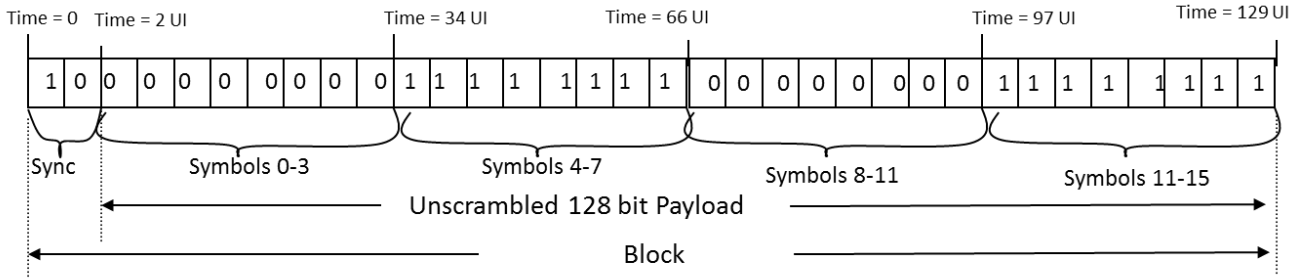
The Evolution of PCI Express® continues with PCIe 5.0 with backwards compatibility

Agenda

- Background
- **128b/130b Changes for PCIe 5.0**
- Transmitter Equalization and Training
- Precoding
- Alternate Protocol Negotiation
- Testability Features
- Configuration Registers for 32GT/s
- Summary

Ordered Set Changes for 32GT/s

- **Electrical Idle Exit Ordered Set Sequence**
  - EIEOS changed to 32 0s followed by 32 1s (same 1G clock pattern as before)
  - 2 back to back EIEOS must be sent (extend the length of clock pattern to be same as Gen 4 for EI exit detection)



(Electrical Idle Exit Ordered Set at 32.0 GT/s data rate)

- **SDS:** Symbols 1-15 changed to 87h (from 55h) to avoid 16G clock pattern prior to data stream
- **SKP OS:** Symbols 0-11 changed to 99h (from AAh) to avoid 16G clock pattern during data stream

TS1/ TS2 Ordered Sets Changes for 32.0 GT/s Data Rate

- **Symbol 4: Data Rate Identifier**
  - Bit 5: 32GT/s data rate supported
- **Symbol 5: Training Control**
  - Bit 7:6: Enhanced Link Behavior Control (only defined for Polling and Configuration when LinkUp = 0b; else Reserved)
    - 00b: Full Equalization required
    - 01b: Equalization bypass to highest data rate (>= 32GT/s) supported
    - 10b: No equalization needed + EQ bypass to highest data rate
    - 11b: Modified TS1/TS2 supported (EQ bypass options advertised there)

  Symbol 5 changes used to negotiate EQ bypass/ modified TS1/TS2 negotiation (for things like alternate protocol negotiation) where as Symbols 6 and 7 changes are for precoding (for avoiding the 16G clock pattern for those receivers sensitive to it)

- **Symbol 6:** EQ TS1/TS2 for 32GT/s rate: Bit 0: “Transmit Precode Request” and bits 2:1 Reserved
- **Symbol 7:**
  - 128b/130b: Bit 6: Transmitter Precoding on when operating at >=32GT/s
  - 128b/130b EQ TS2: Bit 0: Transmitter Precode Request for 32GT/s EQ

Modified TS1/TS2 Ordered Sets

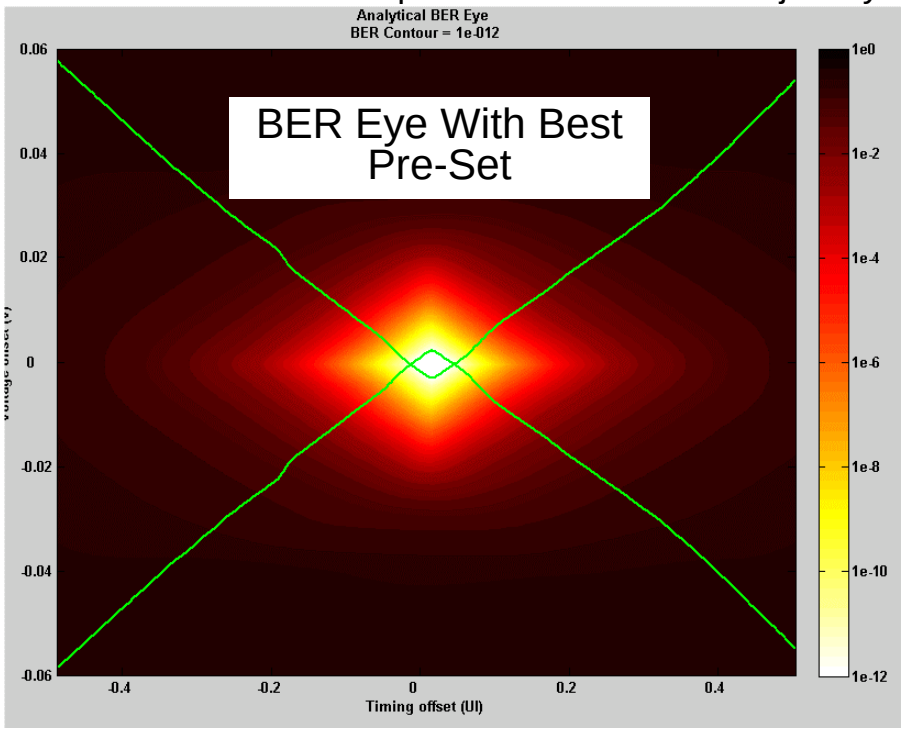
Symbol Number	Description
0	COM (K28.5) for Symbol alignment.
1	Link Number
2	Lane Number within Link : 0-31, PAD. PAD is encoded as K23.7.
3	N_FTS. The number of Fast Training Sequences required by the Receiver: 0-255.
4	Data Rate Identifier: Same as regular TS1/TS2
5	Training/ Equalization Control Bit 0 - Equalization bypass to highest rate support Bit 1 - No Equalization needed. Bit 3:2 - Reserved Bit 4 = 0b, No Retimers present; 1b: One Retimer is present Bit 5 - Two Retimers Present Bit 7:6 = 11b
6	For Modified TS1: TS1 Identifier, encoded as D10.2 For Modified TS2: TS2 Identifier, encoded as D 5.2
7	For Modified TS1: TS1 Identifier, encoded as D10.2. For Modified TS2: TS2 Identifier, encoded as D 5.2
8-9	**Bit 2:0: Modified TS Usage** **000b: PCIe protocol only** **001b: PCIe protocol only with vendor defined Training Set Messages** **010b: Alternate protocol(s)** **011b through 111b: Reserved** **The values advertised in this field must be consistent with the `Modified TS Usage Mode Selected' field of the 32GT/s Control register and the capabilities of the device. These bits are bits[2:0] of Symbol 8.** **Bit 15:3: Modified TS Information 1 if Modified TS Usage = 001b or 101b; else Reserved.**
10-11	**Vendor ID if Modified TS Usage = 001b** **Alternate Protocol ID/ Vendor ID if Modified TS Usage = 010b** **Reserved for other cases**
12 – 14	**If Modifed TS Usage = 001b or 010b, Modified TS Information 2 for the specific usage Else Reserved**
15	**Bit-wise even parity of Symbols 4 through 14.**

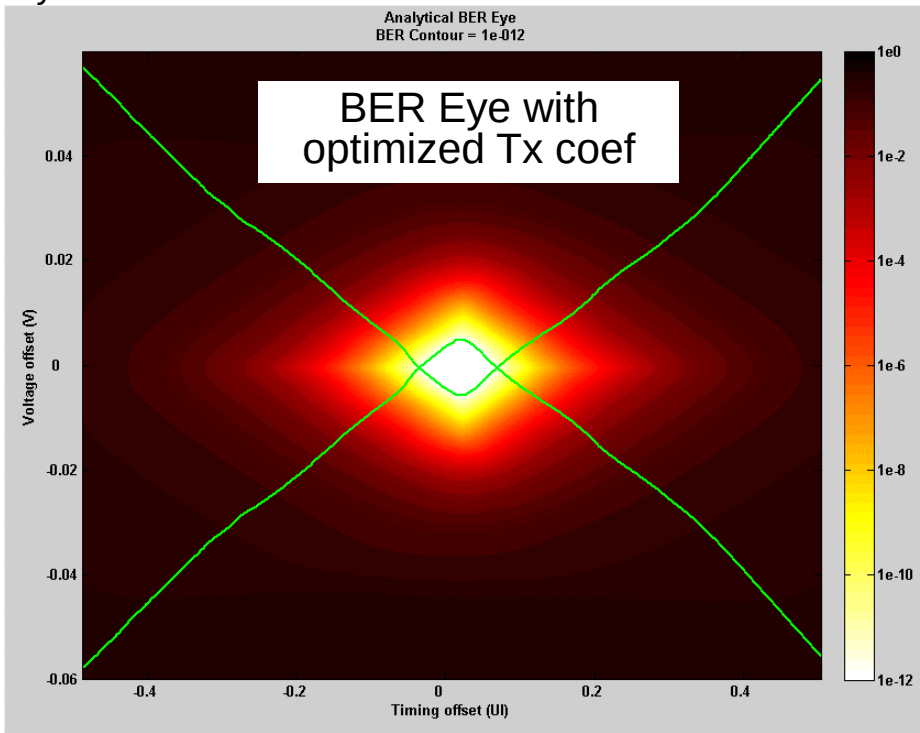
Agenda

- Background
- 128b/130b Changes for PCIe 5.0
- **Transmitter Equalization and Training**
- Precoding
- Alternate Protocol Negotiation
- Testability Features
- Configuration Registers for 32GT/s
- Summary

Transmitter Equalization

- **2.5GT/s and 5GT/s: Fixed de-emphasis for Link**
- **8GT/s, 16GT/s, and 32GT/s: Analysis demonstrates need for per Tx-Rx EQ**
  - Variations in receiver design, channel, PVT
  - Adjust each Tx by its Rx individually
  - Start with a preset value and then adjust dynamically



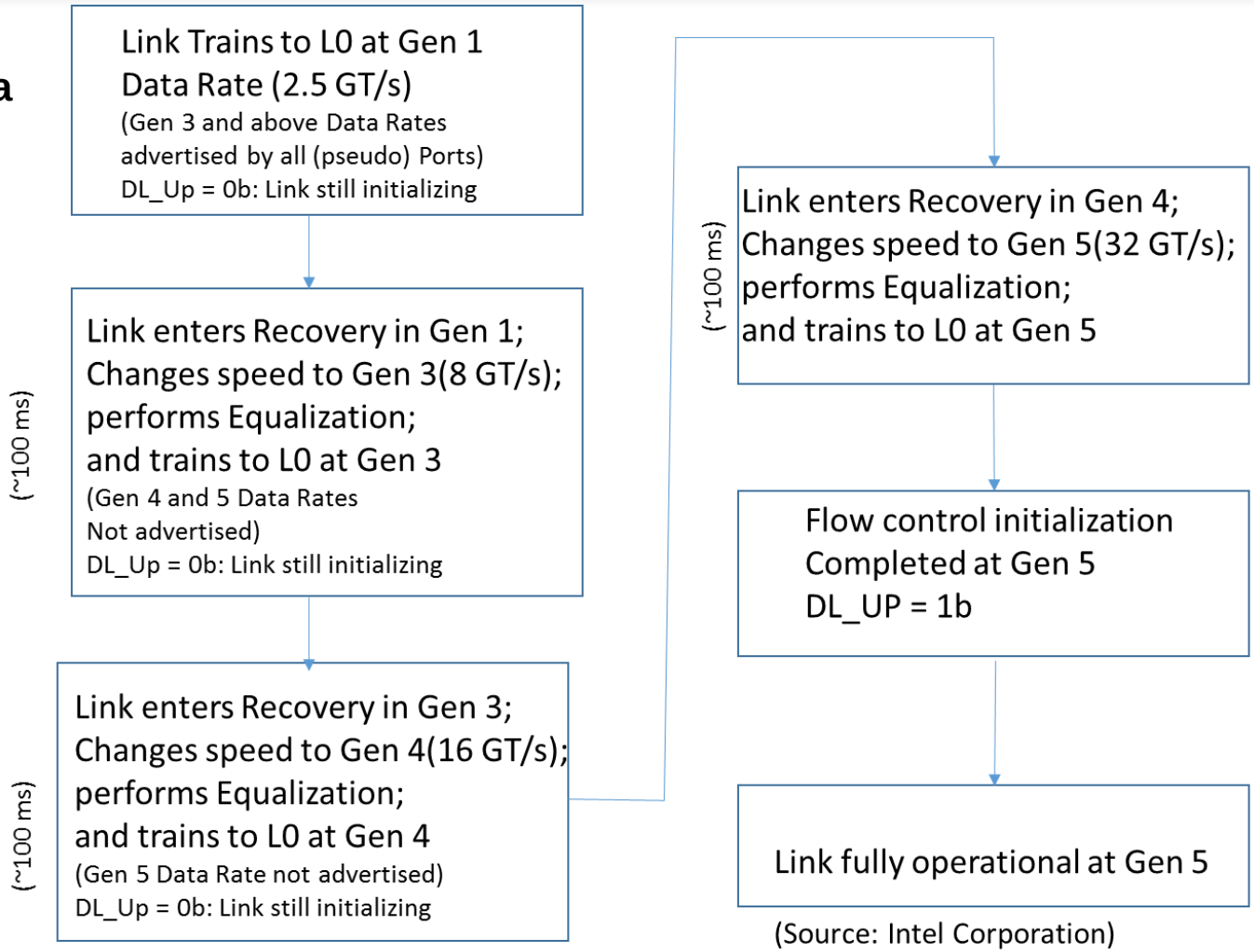


Results from an 18" 2C channel at 8GT/s **Source: Intel Corporation**

Co-efficient based Tx EQ provides better margin

Equalization at 32GT/s: Normal Flow

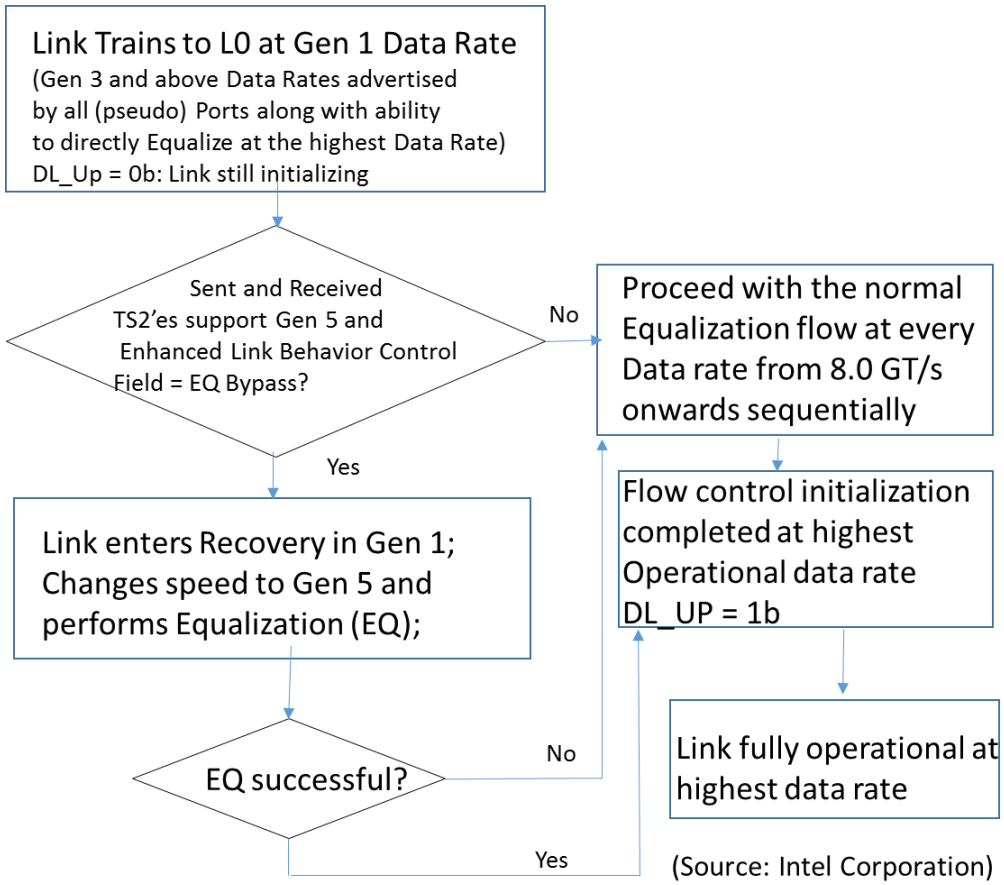
- **DSP withholds advertising higher data rates until EQ has successfully completed at lower data rates**
- **DSP responsible for L0->Recovery transition and advertising 16/32GT/s in the autonomous EQ**
- **No link-width downsizing or power management with autonomous EQ**
- **DLLP exchange withheld till EQ completes for autonomous EQ**



(Source: Intel Corporation)

Equalization Bypass Options

- **Optional feature**
  - All components and Retimers need to support
- **Bypass EQ to 32GT/s**
  - Saves initialization time
  - No EQ in 8GT/s or 16GT/s
  - If Link works in 32GT/s and then has reliability issues, need to perform EQ at 8GT/s or 16GT/s – else needs to run at 5GT/s
- **Option to not perform EQ at all if values from previous EQ can be retrieved from storage**
  - E.g., Link reset, or stored EQ values in persistent storage



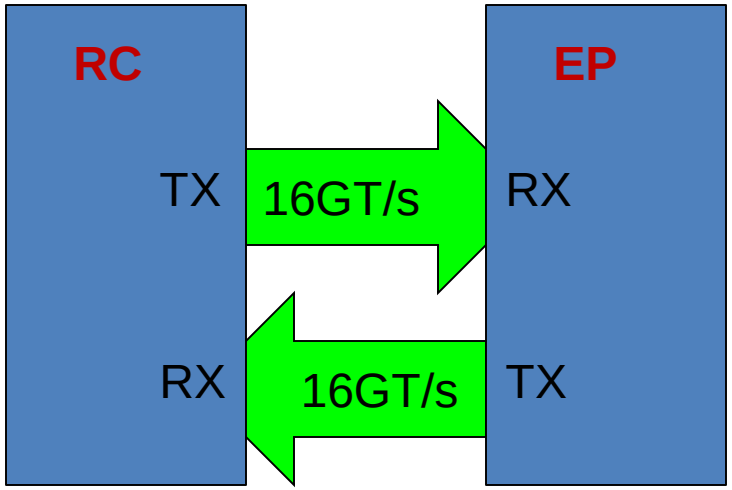
(Source: Intel Corporation)

Equalization – Stage 0

Starting TxEq presets sent from Downstream Port to Upstream Port on a per Lane basis using 128b/130b EQ TS2 (prior to Link going to 32GT/s) (optionally USP can request DSP also)

- Preset is transferred in EQ TS2 Ordered Sets
  - ✓ A Port may use a different preset in its Tx than it requests its Link Partner to use
  - ✓ Preset values (for both ports) come from the Downstream Port's (HwInit) CSR (USP's request, if implemented, is implementation specific)

Preset
0
1
2
3
4
7
8
9
10

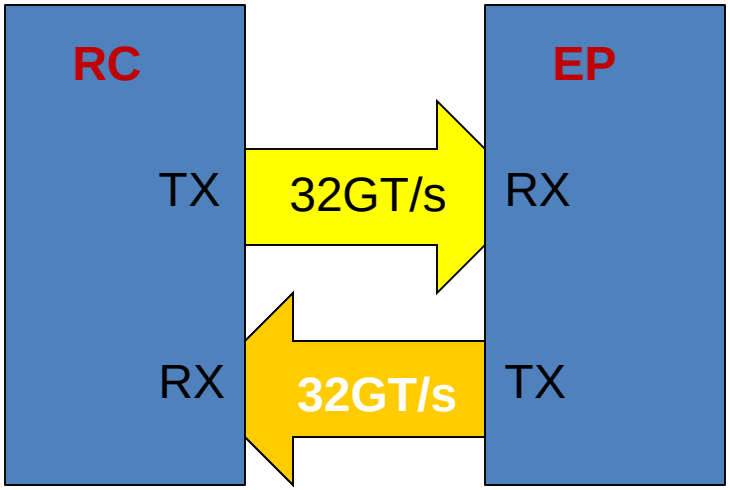


Source: Intel Corporation

Equalization – Stage 1

Starts after Link transitions to 32GT/s. Both Ports use the TxEq Presets from Stage 0. Corresponds to Phase 1 in Downstream Port and Phases 0 and 1 in Upstream Port.

Expectation is that link will operate “good” enough to allow progression at 32GT/s (BER <= 10^{-4}) in 24 msec; else link will go to a lower Data Rate



Source: Intel Corporation

Back Channel – Stages 2/3

Stage 2: Intended for Upstream Port to achieve BER <= 10^{-12}. Starts at the preset. Coefficients/ presets are exchanged in sub-loops until this is accomplished within 24 ms A Port may decide not to make any new requests. Corresponds to Phase 2

- Receiver full swing (FS) defines granularity of coeff
  - ✓ Table at bottom-right is for illustrative purposes
  - ✓ X-axis is pre-cursor, y-axis post-cursor, diagonal defines the boostline
  - ✓ Each tile represents a coeff (e.g. p7=4/6, p8=5/5, etc)
  - ✓ Numbers in tiles represent presets; black tiles are illegal coeff space

Example: start from preset 7 (coef=4/6)

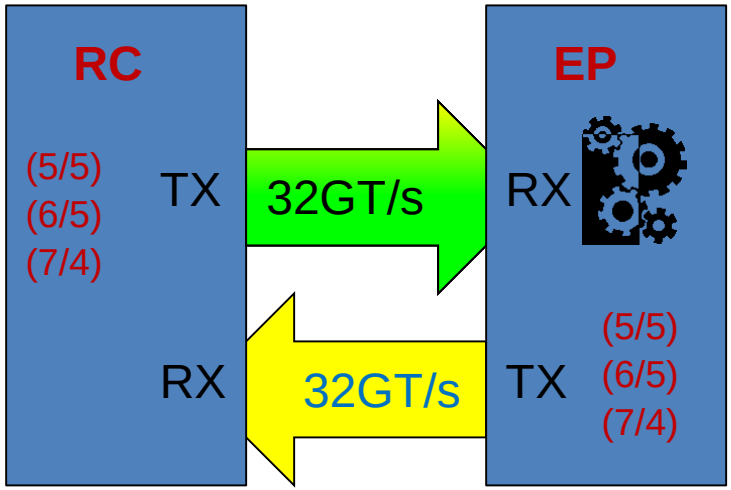
1st sub-loop

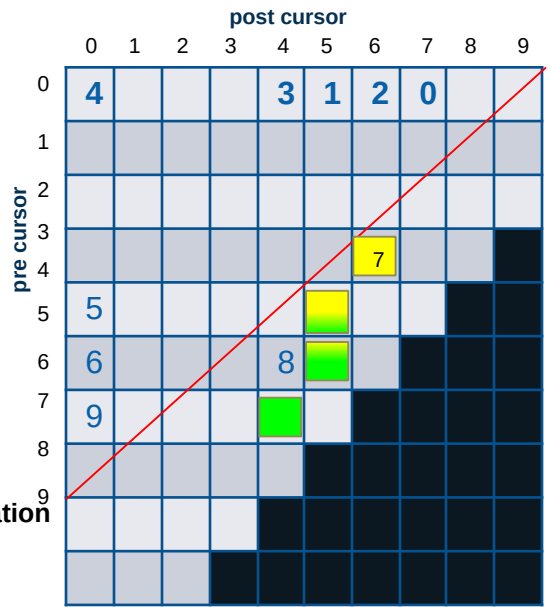
a. EP Rx eval reveals need for less post, more pre
b. EP sends (5/5) to RC
c. RC applies (5/5) to TX
d. RC echo's (5/5) to EP

2nd sub-loop

a. EP Rx eval needs more pre, post ok
b.→ d. repeat with (6/5)

3rd sub-loop finds good result with (7/4) so moves to phase 3





Source: Intel Corporation

Stage 3/ Phase 3 is same as phase 2 in opposite direction
Downstream Port may skip Phase 2/ 3 if presets are good enough for Link
Retimers perform independent EQ but within the Phases of the Ports
Retimer Enhancements for 32GT/s : "Retimer Equalization Extend" to complete Ph 2/ 3

Equalization Procedure

- **Expected to be done once autonomously after Link trains to L0**
  - No DLLP/TLP exchange till equalization completes
  - Ensures no TLP timeout as equalization can take more than 100 ms
  - Software polls DL_Active prior to accessing downstream component
- **Software can perform EQ by accessing CSRs in Downstream Port**
  - Must ensure no side-effects (e.g., no timeout)
- **A device may withhold 8GT/s or 16GT/s or 32GT/s Data Rate (and EQ)**
  - If its associated software can guarantee no side effects of doing equalization when it advertises 8GT/s or 16GT/s or 32GT/s Data Rate
- **Error during equalization or later**
  - Not expected to redo equalization except error condition
  - Downstream Port can redo EQ
  - Upstream Port must report in its register and request
  - Downstream Port has two choices: (i) redo equalization (ii) log and report

Agenda

- Background
- 128b/130b Changes for PCIe 5.0
- Transmitter Equalization and Training
- **Precoding**
- Alternate Protocol Negotiation
- Testability Features
- Configuration Registers for 32GT/s
- Summary

Precoding – The Problem Statement

- **For certain DFE settings (high H1/H0), possibility of contiguous burst errors, with certain patterns like the clock pattern (e.g., SKP)**
  - Potential problem if the errors cause aliasing with LCRC / CRC – need to protect data stream in that case
  - Pre-coding XORs the current bit with previous bit:
    - Pros: Converts a long contiguous burst error into two random bit flips => CRC detects
    - Cons: Converts single bit flip to 2 bit flips

Incoming Pattern:	0 1 0 1 0 1 0 1 0 1 0 1 0 0…
Error bit on wire:	0 1 0 0 0 0 0 0 0 0 0 0 0 0…
Error after DFE:	0 1 1 1 1 1 1 1 1 1 1 1 1 0..
Error w/ Precoding:	0 1 0 0 0 0 0 0 0 0 0 0 0 1..

(a: Contiguous error burst reduced to two bit flips with precoding)

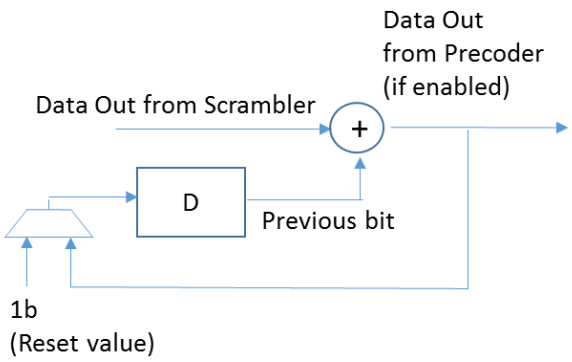
Incoming Pattern:	0 1 1 1 0 0 1 0 1 0 1 0 1 1 …
Error bit on wire:	0 1 0 0 0 0 0 0 0 0 0 0 0 0 …
Error after DFE:	0 1 0 0 0 0 0 0 0 0 0 0 0 0 ..
Error w/ Precoding:	0 1 1 0 0 0 0 0 0 0 0 0 0 0..

(b: Single bit flip increases to two bit flips with precoding)

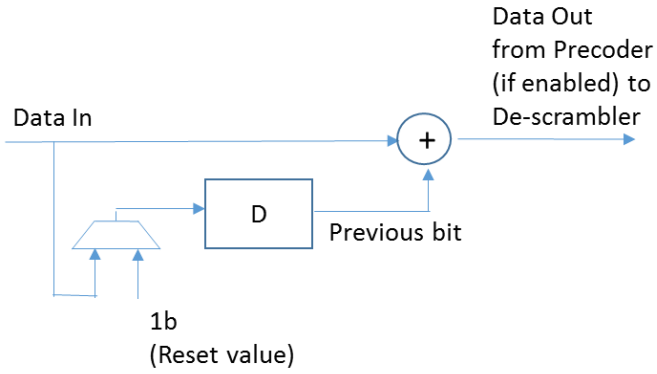
- **Constraints: no fundamental change to 128b/130b encoding (e.g., no scrambling of EIEOS, SKP OS, etc)**

Precoding Mechanism

- **Requested by Receiver prior to entering the 32G data rate through EQ TS2 or 128b/130b EQ TS2 Ordered Sets and the precoding mode stays on through the entire 32G data rate till the next equalization**
  - Transmitter must pre-code the scrambled bits in 32G data rate when requested
  - All Lanes in each segment must request the same way (either with precoding on or off)
- **Only scrambled bits are pre-coded when it is on. Precoding advances with scrambler. Reset at block boundary with "1"**
- **When precoding is on for Lane 0 of Transmitter, the "Transmitter Precoding On" bit of the 32GT/s Status Register is set to 1b; else 0b**
- **When a Receiver requests precoding: the "Precoding Requested" bit in the 32GT/s Status Register is set to 1b; else 0b**



(a. Precoding on Tx side)

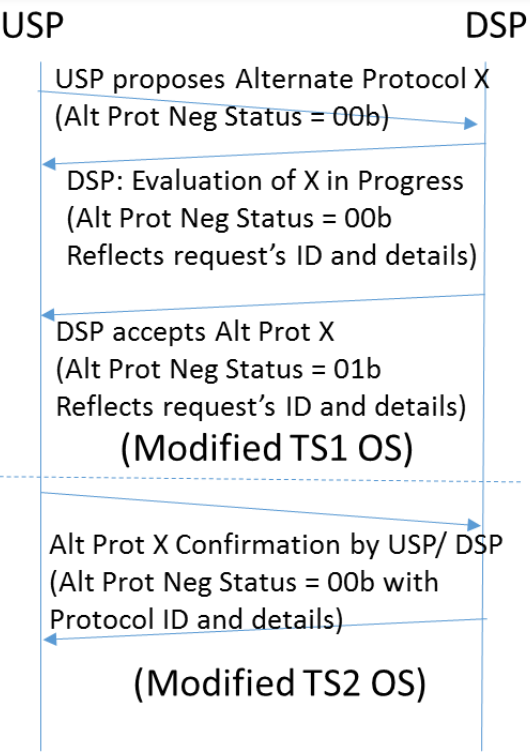


(b. Precoding on Rx side)

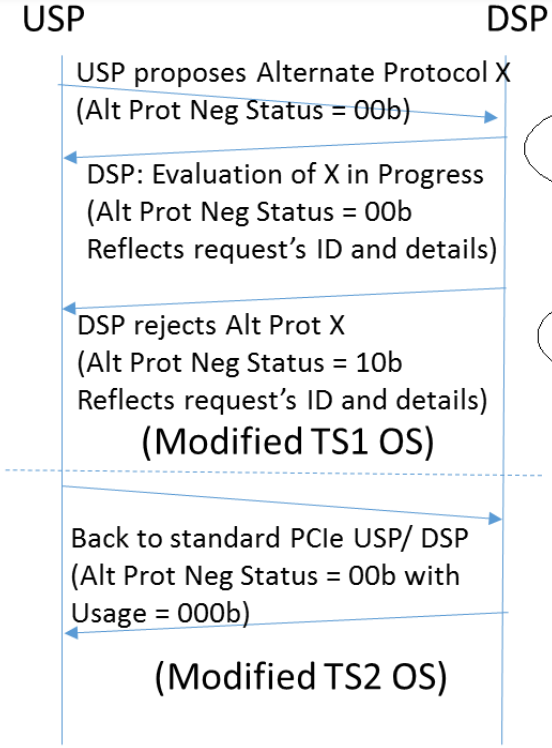
Agenda

- Background
- 128b/130b Changes for PCIe 5.0
- Transmitter Equalization and Training
- Precoding
- **Alternate Protocol Negotiation**
- Testability Features
- Configuration Registers for 32GT/s
- Summary

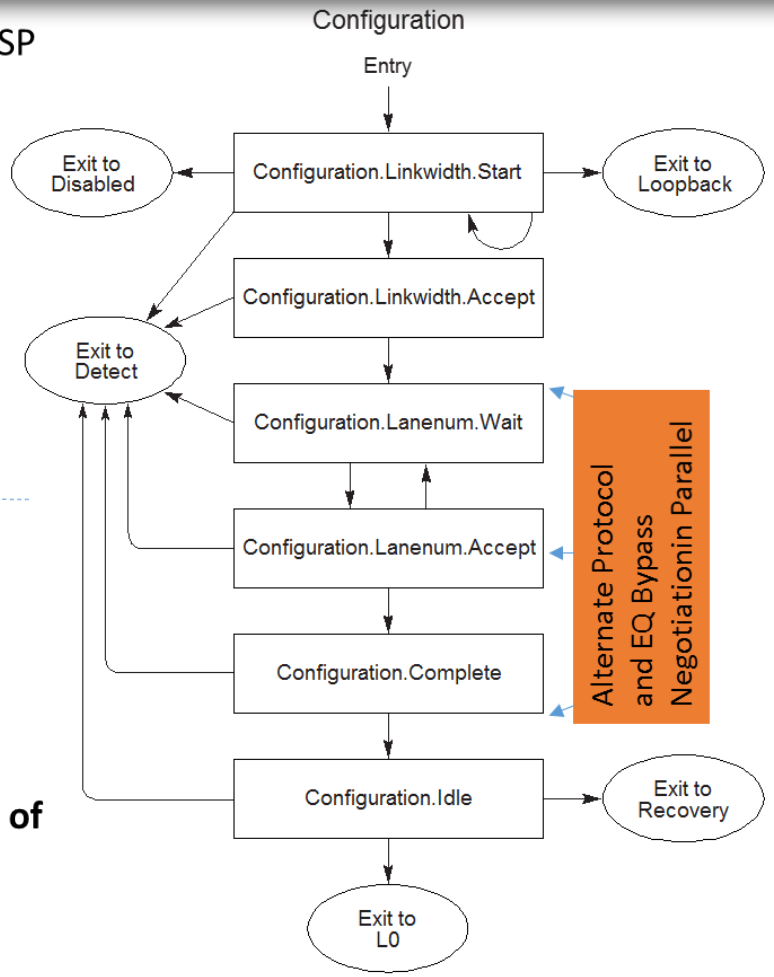
Alternate Protocol and Vendor Defined Training Messages Negotiation



(a: Successful negotiation of Alternate Protocol X)



(b: Unsuccessful negotiation of Alternate Protocol X results in PCIe being selected)



(Modified Training Sets in Config State of LTSSM to negotiate Protocol)

[Alternate Protocol Information 1: Bits 3:4 – alt prot neg status, Bits 5:15: Alternate protocol details}

(Source: Intel Corporation)

Agenda

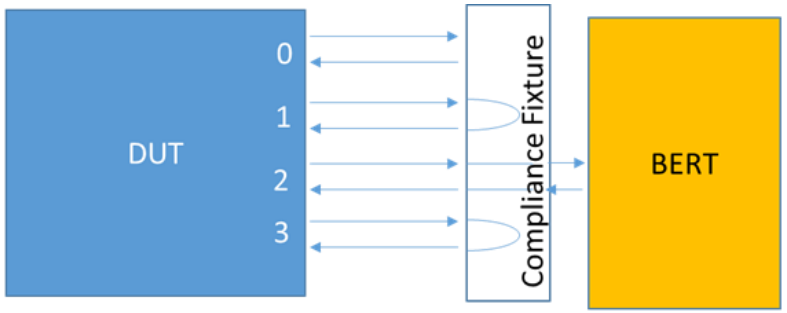
- Background
- 128b/130b Changes for PCIe 5.0
- Transmitter Equalization and Training
- Precoding
- Alternate Protocol Negotiation
- **Testability Features**
- Configuration Registers for 32GT/s
- Summary

Testability

- **<u>New Mandatory Feature with PCIe 5.0</u>: Rx testing with BERT so far has been on a single Lane basis, restricted to Lane 0 only**
  - BERT trains the DUT to L0 (hence the Lane 0 dependency), performs EQ through Recovery, transitions from Recovery to Loopback (with BERT as master), and then tests the receiver
- **Issue: We expect the cross-talk contributions from the package to be a significant contributor to Rx Margin loss at 32GT/s**
  - Multi-Lane BERTs too expensive
- <u>Other</u>: Compliance and Modified Compliance patterns the same as PCIe 4.0 except EIEOSQ causes additional OS(es)
  - Mechanisms for entry / exit w/ CLB/CBB, CSR-based or TS-Ordered Set based the same

Solution: A mechanism to enable testing of any Lane with BERT while the rest of the Lanes will transmit (and optionally receive) the modified compliance pattern to mimic the cross-talk behavior (far-end – FEXT or near-end NEXT)

Applicable only for 32GT/s (and higher) data rate(s)

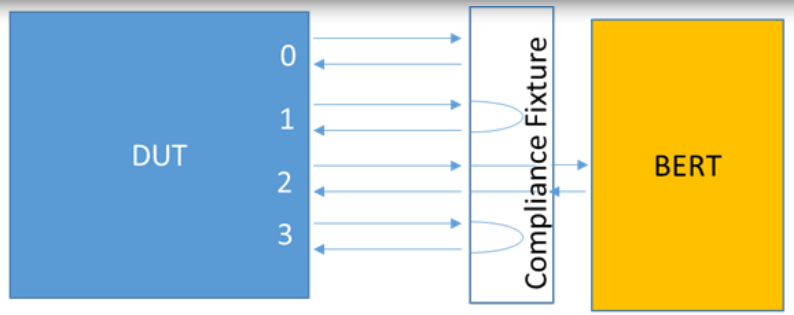


(BERT connected to Lane 2 while Lanes 1 and 3 are looped back to mimic FEXT but Lane 0 is not looped back to mimic NEXT)

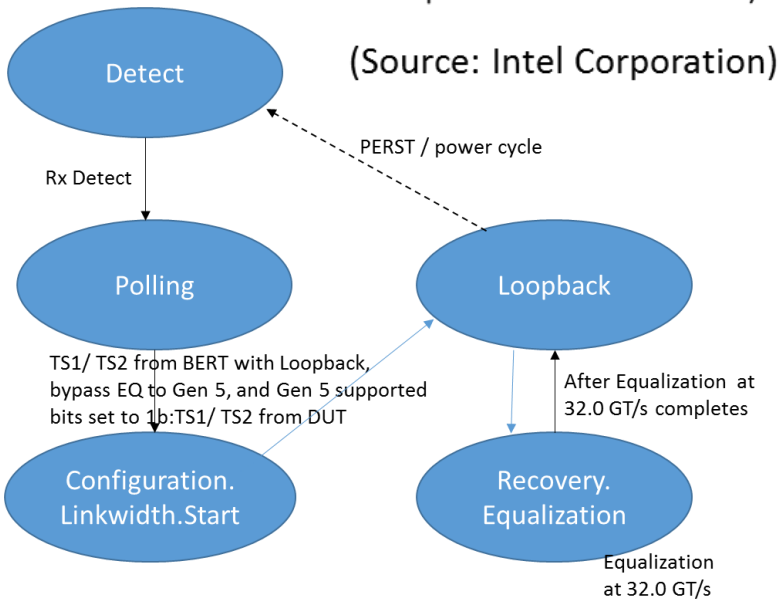
(Source: Intel Corporation)

LTSSM Changes to Allow Rx Testing of any Lane by a BERT with FEXT/ NEXT Effects

- All / Some Lanes will detect a Receiver in Detect (depending on FEXT/ NEXT set up in fixture)
- BERT sends TS1/ TS2 Ordered Sets through Polling.Active, Polling.Configuration, and Configuration.Linkwidth.Start with Loopback bit, Enhanced Link Behavior Control bits set to 01b, and PCIe 5.0 data rate supported bits set to 1b
- All Lanes that detected a Receiver train with their self-looped-back TS1/ TS2 Ordered Sets where BERT is not connected
- LTSSM moves from Configuration.Linkwidth.Start to Loopback.Entry - the Lane number assignment would be the default Lane number
- LTSSM moves to Recovery.Equalization and completes the equalization flow at 32GT/s data rate (skip over 8GT/s and 16GT/s data rate – in this test mode it is mandatory for the DUT to skip over the equalization to the highest data rate)
- At the conclusion of Equalization, LTSSM moves back to Loopback.Entry
- BERT is the loopback master and runs the receiver margining tests



(BERT connected to Lane 2 while Lanes 1 and 3 are looped back to mimic FEXT but Lane 0 is not looped back to mimic NEXT)



(LTSSM transitions for the per-Lane Rx testing with FEXT/NEXT)

(Source: Intel Corporation)

Agenda

- Background
- 128b/130b Changes for PCIe 5.0
- Transmitter Equalization and Training
- Precoding
- Alternate Protocol Negotiation
- Testability Features
- **Configuration Registers for 32GT/s**
- Summary

32GT/s Related Configuration Registers

- **32GT/s Data Rate reflected in existing registers**
  - E.g., `Supported Link Speeds Vector' in `Link Capabilities 2 Register'
- **Physical Layer 32GT/s Extended Capability structure**
  - Capability reflects: EQ bypass, no EQ needed, modified TS support modes
  - Control provides ability to disable the capabilities above (e.g., EQ bypass disable)
  - Status reflects the Equalization status (success of various phases) as well as precode requested, precode on, modified TS received, EQ bypass capability advertised by Link partner
  - Information from modified TS, if received, including status in 2 data registers
  - Information on modified TS, if transmitted, in 2 data registers
  - Per- Lane Equalization Control bits similar to prior speeds
- **Note: Local Data Parity Mismatch (10h) (one bit per Lane) as well as First and Second Retimer Parity Mismatch (14h and 18h) NOT in this structure but in the 16GT/s Extended Capability structure. Margining also reuses the 16GT/s Margining Extended Capability structure**

Agenda

- Background
- 128b/130b Changes for PCIe 5.0
- Transmitter Equalization and Training
- Precoding
- Alternate Protocol Negotiation
- Testability Features
- Configuration Registers for 32GT/s
- **Summary**

Summary

- **PCIe 5.0 Base Spec mature at Rev 0.7 level**
- **Introduces a new Data Rate at 32GT/s**
  - Doubled bandwidth again – now at 5th generation!
  - Expect Retimers for longer channel reach
- **Leverage existing Encoding, Tx Equalization, and compliance mechanisms**
- **Enhancements include better testability features, speeding up link training time (bypassing EQ), and support for alternate protocols**
- **Track the PCIe 5.0 Base Spec and provide your feedback**

Thank you for attending the PCI-SIG Developers Conference 2018.

For more information, please go to www.pcisig.com

Don't forget to submit your feedback via the mobile app!

Download the **Crowd Compass** app and then search for **PCI-SIG Developers Conference** or entering the following URL into your mobile browser: https://crowd.cc/s/1rKy0
Enter event code: **DevCon2018**

Alternatively, access here: https://crowd.cc/pcisig2018

Note: Create an account within the app so Admin knows who to contact if selected as the prize winner.

Each session feedback is provided is equivalent to 1 raffle entry (up to 11 sessions).
General survey feedback = 1 raffle entry.